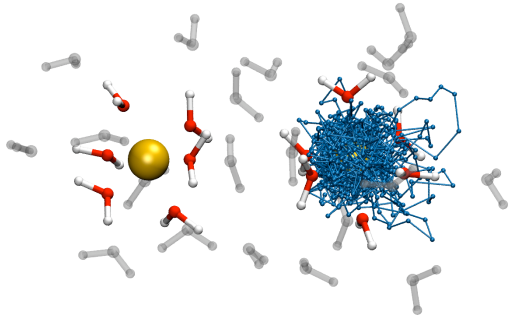
BES-NERSC Workshop
Washington DC
9-10 Feb. 2010



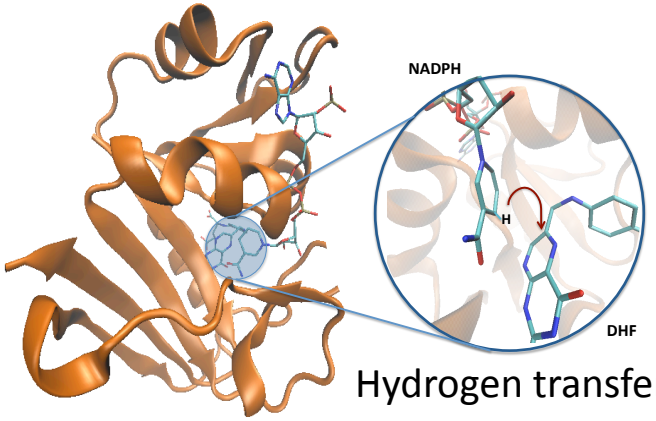
Reaction Dynamics in Complex Molecular Systems

Thomas Miller
California Institute of Technology

Project Overview

Summary:

Develop computational methods to understand complex dynamics in chemical and biomolecular systems.

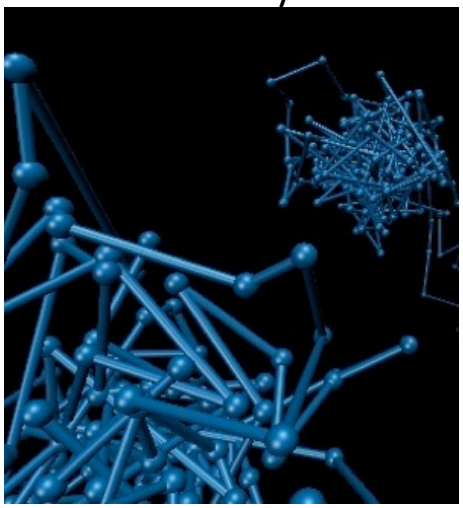


Hydrogen transfer

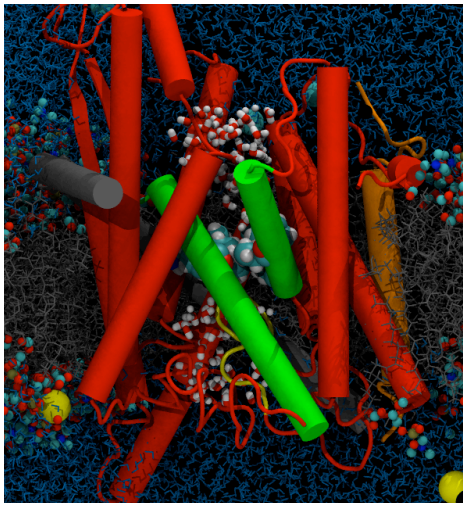
Computational methods:

1. Classical molecular dynamics
2. Semiclassical and path integral methods for quantum effects in dynamics and statistics
3. Rare event sampling methods for determining reaction mechanisms and kinetics
4. Free energy sampling and coarse-graining methods for studying conformational landscapes
5. Force fields and electronic structure theory methods for calculating on-the-fly potentials

Non-adiabatic dynamics

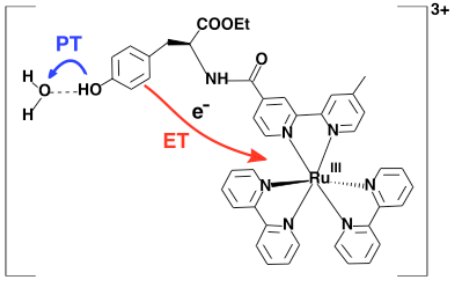


Protein Translocation

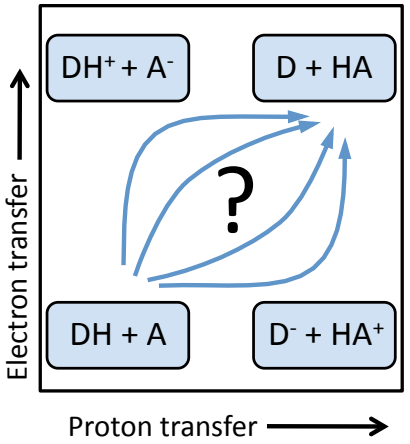


Scientific objectives in next 3-5 years

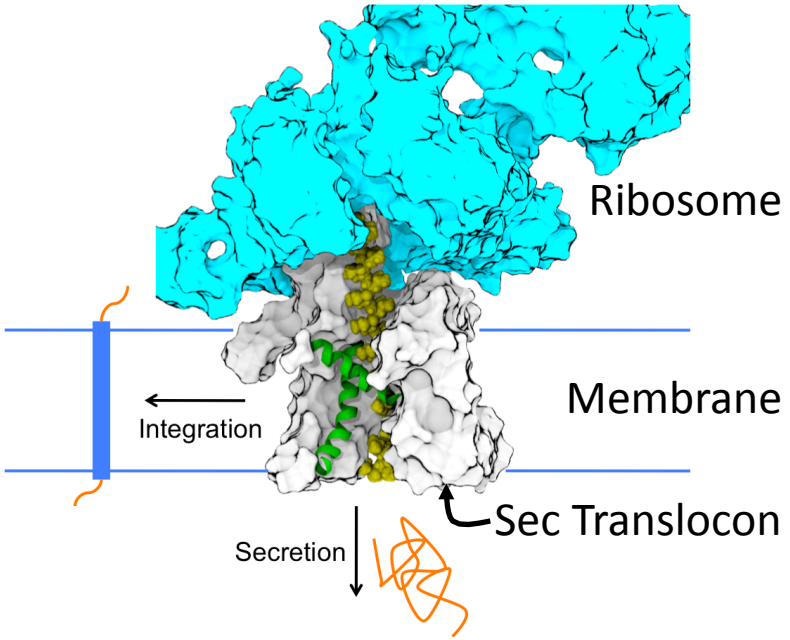
- Reaction dynamics involving coupled electronic and nuclear dynamics
  - Proton-coupled electron transfer reactions in photocatalysts and hydrogenases
  - Reactions of the solvated electron



Photosystem II

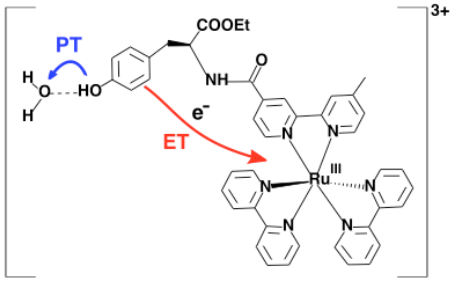


- Protein transport across cellular membranes
  - Regulation of the secretion vs. integration pathways
  - Protein-protein interactions and Substrate-induced conformational switching
  - Non-equilibrium and glassy dynamics

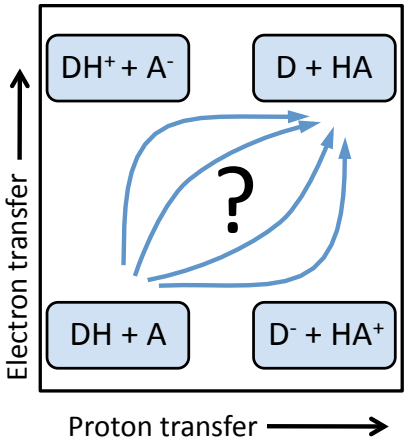


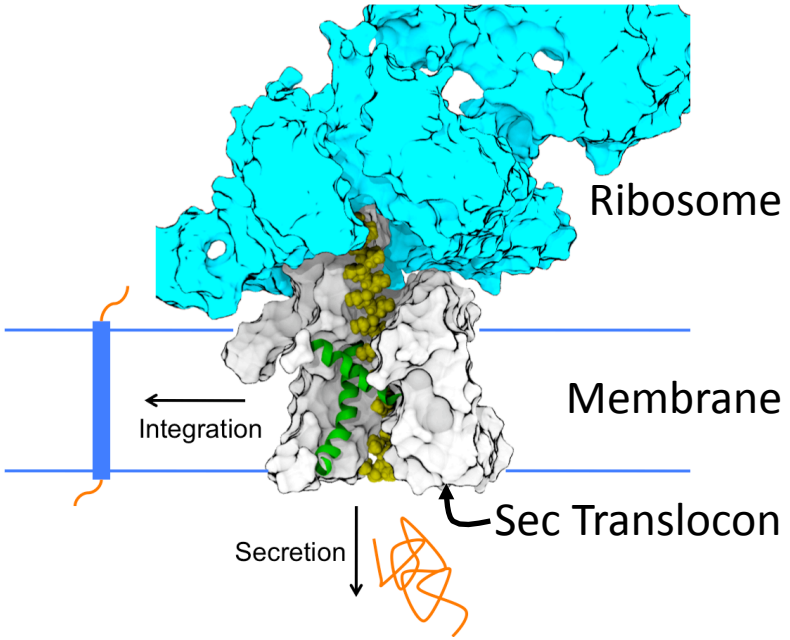
Challenges for computation

- Bridging length-scales and time-scales in molecular simulation
  - Electron transfer (sub-fs)
  - Proton transfer (fs)
  - Solvent fluctuations and side-chain dynamics (ps-ns)
  - Conformational dynamics (ns-ms)
  - Reaction waiting times (ns-s)
  - ATP hydrolysis in molecular machines (s)
- Simulating dynamics of coupled quantum mechanical and classical systems
- Calculating reaction mechanisms and rates without transition state theory assumptions



Photosystem II

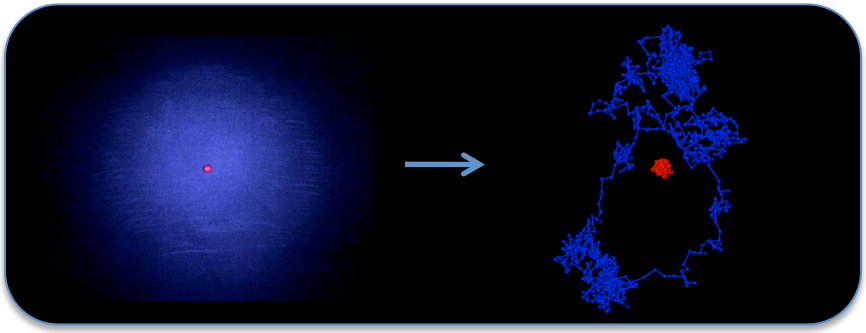




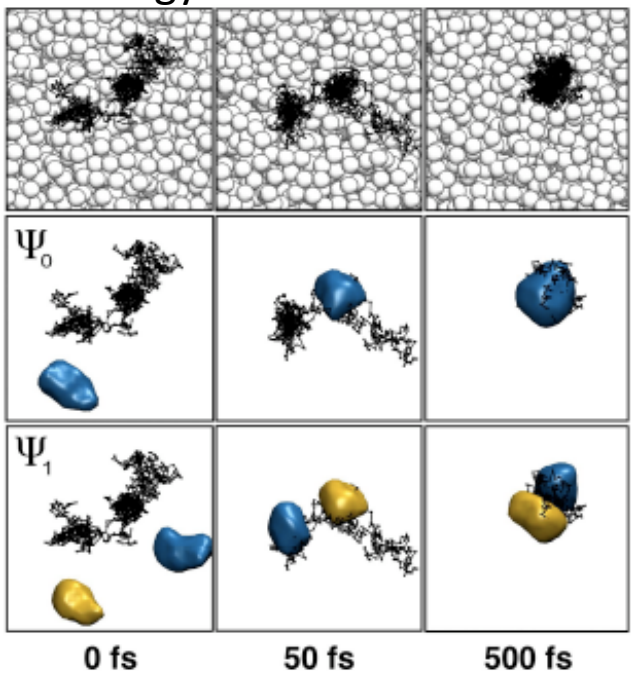
Unifying theme: Classical MD Trajectories

- Including quantum effects with Path Integral Molecular Dynamics
  - Exact quantum statistics
  - Semiclassical (approximate) dynamics
    - Ring polymer Molecular Dynamics
    - Centroid Molecular Dynamics

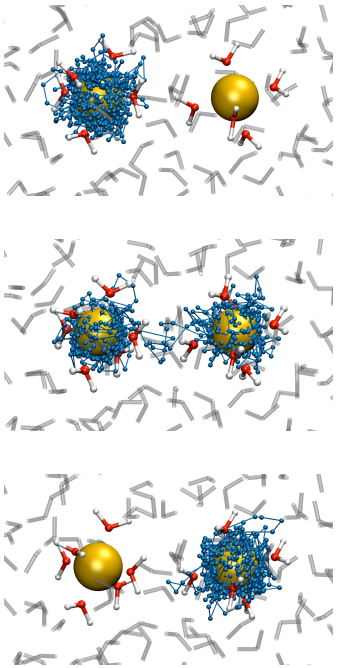
Path integral classical isomorphism



Localization Dynamics of High Energy Electrons in Water

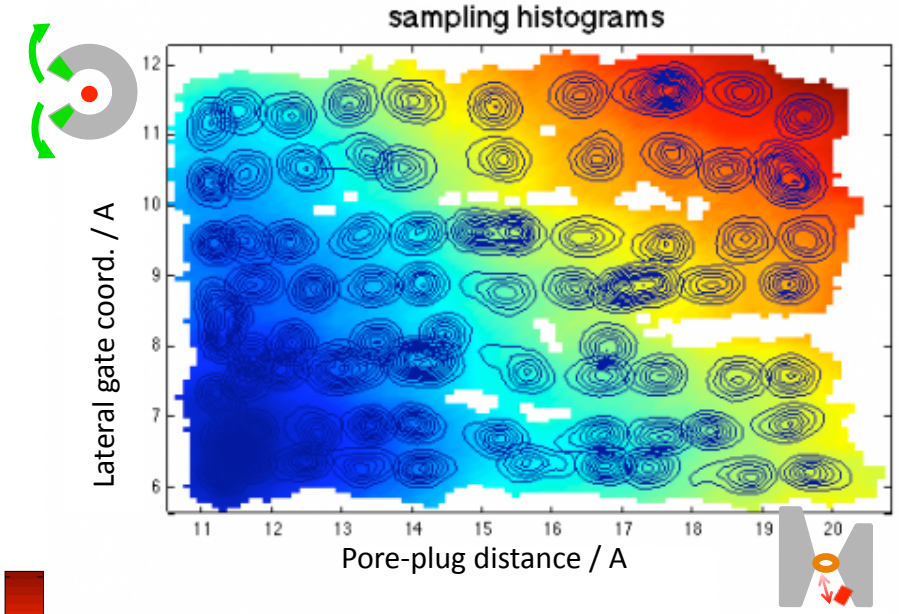


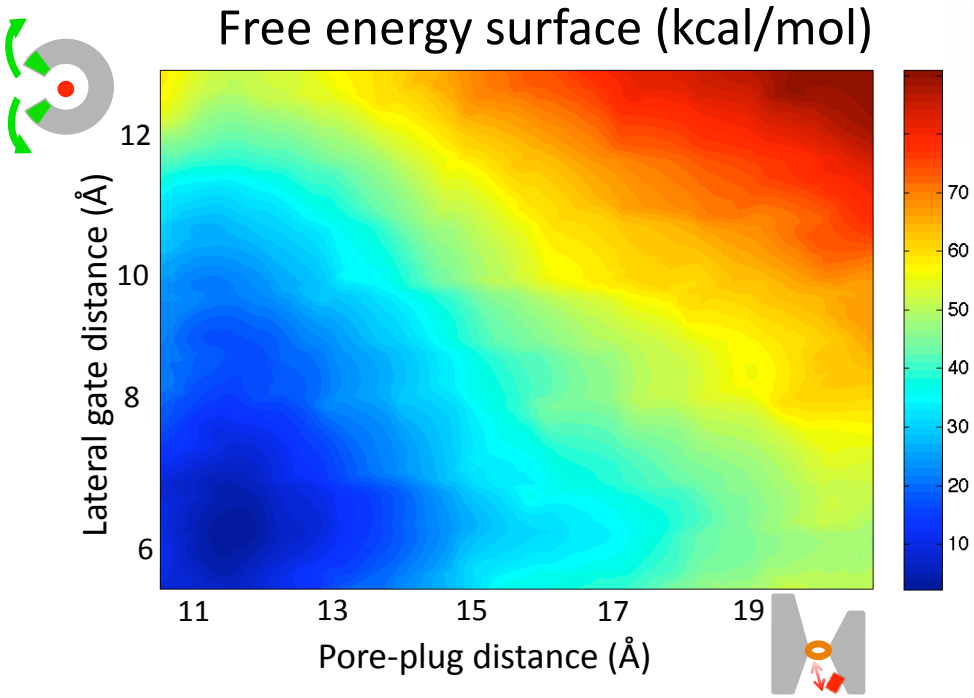
Electron Transfer Dynamics

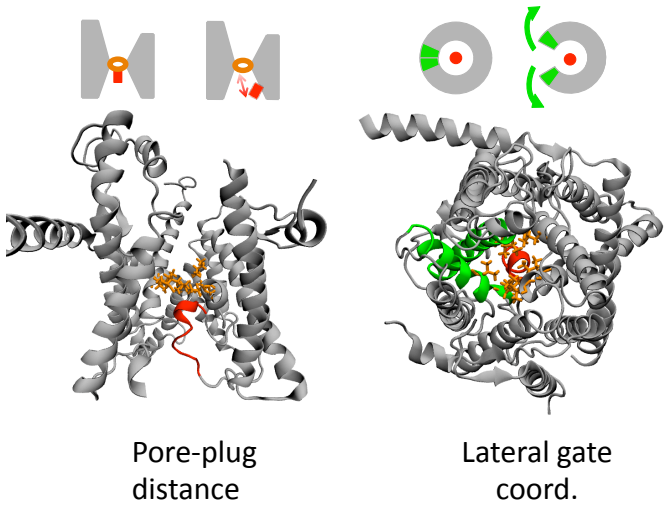


Unifying theme: Classical MD Trajectories

- Free energy sampling methods for the mapping conformational landscapes
  - WHAM Algorithm
  - Replica Exchange Methods

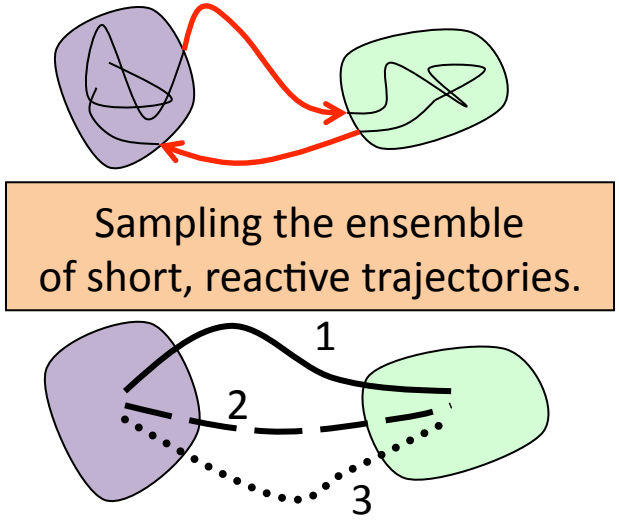




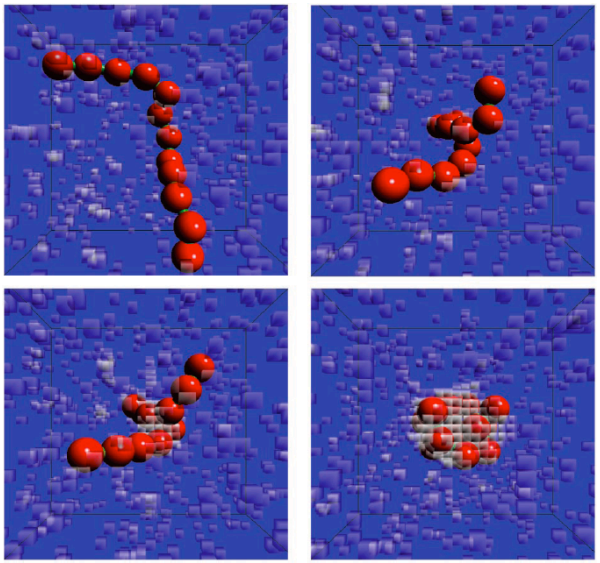


Unifying theme: Classical MD Trajectories

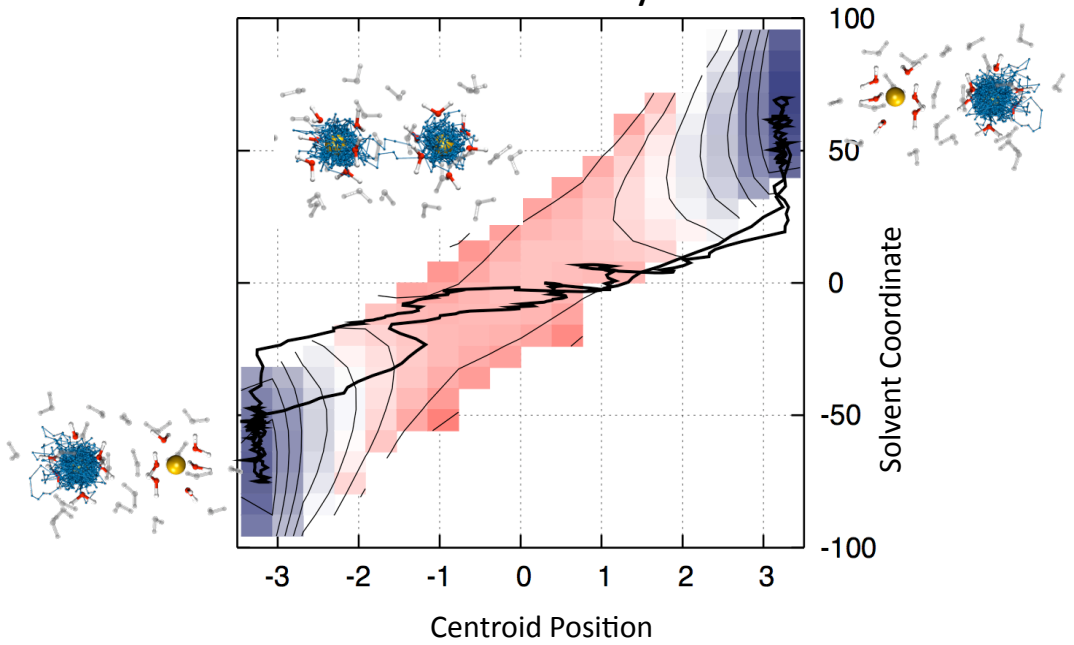
- Calculating reaction mechanisms and rates without transition state theory assumptions
  - Rare-event sampling Methods
  - Transition Path Sampling
  - String Method in Collective Variables
  - Forward-Flux Sampling



Solvent Dynamics
During Polymer Collapse



Electron transfer dynamics



2. Current HPC Requirements

- Machines used
  - Franklin, Jacquard, Jaguar (NCCS)
  - ~1.5 M hours used last year, ~4 M expected use this year
- # cores, amount of memory, input/output, disk storage typically used
  - 32-128 cores per trajectory. Bundling of trajectories is trivially parallel.
  - Low demands in terms of memory, I/O, and disk storage.
- Necessary software, services or infrastructure
  - MPI, FFTW
- Data transfer requirements (within NERSC or to/from NERSC)
  - 20 GB/week
- Current primary codes and their methods or algorithms
  - MD Codes: Modified versions of NAMD, DLPOLY, GROMACS
- Known limitations / obstacles / bottlenecks
  - Parallelization of the force evaluated is limited communication speed/latency.
- # of runs, types, length, reasons
  - Hundreds or thousands of relatively trajectories needed.
  - Queue time limits are not a major hindrance for these calculations.

3. HPC Usage and Methods in 5 Years

- Upcoming changes to codes/methods/approaches
  - Increased use of on-the-fly electronic structure methods: plane-wave, DVR-basis, and embedded DFT strategies
- Estimate of MPP hours needed to achieve science goals
  - 30 M
- Estimate of #cores, memory, storage, I/O required
  - 400-10000 cores per trajectory
- Changes to necessary software, services or infrastructure
  - Increased use of electronic structure theory codes: Molpro, DFT codes
- Anticipated limitations/obstacles/bottlenecks on 10K-1,000K core system.
  - Force evaluations and individual electronic structure theory calculations increasingly scale to 250-1000 cores. If 1000 trajectories are needed, no obstacles encountered.
- Preparations for, or use of, emerging HPC Architectures and Programming Models
  - MD methods are proving amenable to GPU technologies, etc.

4. Additional Questions

- What significant scientific progress could you achieve with access to 50X NERSC resources now?
  - The accuracy of ab initio potentials is essential for simulations to go from the "understanding" phase to the "predicting and designing" phase. With 50X resources, the methods described could employ those methods, enabling the rational design of synthetic photo-catalysts, electrolyte additives to prevent dendrite formation in lithium batteries, etc.

- Recommendations on NERSC architecture, system configuration and associated service requirements needed for your science.
  - Fast and available cycles, low-latency communication for parallelizability of potential evaluations, large numbers of cores

- What "expanded HPC resources" are important for your project?
  - More CPU hours, faster access to the queues